

DOCUMENT RESUME

ED 474 990

EA 032 425

AUTHOR Fidler, Brian
TITLE External Evaluation and Inspection.
PUB DATE 2002-00-00
NOTE 24p.; In: The Principles and Practice of Educational Management; see EA 032 295.
PUB TYPE Guides - Non-Classroom (055)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS *Accountability; Cultural Context; *Educational Administration; *Educational Assessment; Educational Practices; Elementary Secondary Education; *Evaluation Utilization; Higher Education
IDENTIFIERS Australia (Victoria); England

ABSTRACT

This document is a chapter in "The Principles and Practice of Educational Management," which aims to provide a systematic and analytical introduction to the study of educational management. The structure of the book reflects the main substantive areas of educational leadership and management, and most of the major themes are covered in the volume's 19 chapters, of which this is one. This chapter examines the reasons for increased interest in accountability as it highlights the tension between evaluation for accountability and evaluation for improvement. It details different approaches to evaluation and discusses their contributions to evaluation at system and institutional levels. The reasons for increased interest in accountability are many and varied, ranging from economic competitiveness, to financial stringency, to equality of opportunity, to decentralization. Institutional evaluation, like staff appraisal, is generally held to have two main purposes: accountability and development. Although these two ideas are linked, combining them in the same process has its tensions. Accountability is intended to be made public, but this is inimical to an honest assessment of performance, particularly when it is carried out in a competitive situation. Although accountability is the most often cited reason for external evaluation, this often means little more than the assessment of institutional performance, whereas evaluation is intended to report on the current performance of individuals and institutions as a spur to improvement. Some examples of school inspections in England under the Office for Standards in Education, as well as an overview of school review in Victoria, Australia, are provided. In any kind of evaluation, the criteria for evaluation must be clear so that the evidence base and the basis for judgments are known to the inspectors and the inspected. (Contains 35 references.) (RJM)

ED 474 990

External Evaluation and Inspection.

Brian Fidler

2002

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN
MICROFICHE ONLY HAS BEEN GRANTED BY

N. Swartz

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

2B

EA 032 425

BEST COPY AVAILABLE

External Evaluation and Inspection

Brian Fidler

Introduction

There has been a trend to greater accountability in most educational systems and at most levels and stages within those systems. This has been a spur to the development of evaluative practices. The main focus of accountability has been of giving an account to those outside the educational system. This has implications for the evaluative processes which provide that account.

The chapter examines the reasons for this increased interest in accountability and highlights the tensions between evaluation for accountability and evaluation for improvement. It details different approaches to evaluation and examines their contributions to evaluation at system and institutional levels. Finally, it analyses two different recently designed approaches to school evaluation. One highlights evaluation for system and institutional accountability by inspecting all schools, whilst the other emphasises institutional development and review.

Reasons for accountability

The reasons for increased attention to accountability are many and varied and undoubtedly there has been much 'policy borrowing' by some systems from those that were first in the field. The range of reasons which have been adduced are:

- Economic competitiveness. Nations have been increasingly concerned about their international competitiveness in business and commerce and they have regarded the performance of the educational system as a way of preparing students to become economically productive. This effect has been particularly marked in countries where economic performance has been a cause for concern.
- Financial stringency. Meeting the costs of providing public services has become an issue in all countries. Whilst one element has been the overall proportion of national production consumed by public expenditure, an

equally important effect has been the distribution of spending. There have been rising claims from services such as care for the elderly and health. This has led to pressure on educational spending and a concern to obtain value for money for whatever is spent on the educational system.

- Equality of opportunity. This has three aspects. In addition to the general social and political concern that all citizens in a democracy receive equal access to public services there has been increasing recognition that certain groups either have not profited from their educational opportunities or have dropped out. These groups have particularly included the socially disadvantaged and ethnic minorities. This is an especially insidious effect as the proportions of these groups in education decrease in the more advanced sectors, particularly higher education. Two consequences of this have been recognised. Without educational qualifications such groups are unable to escape their social deprivation as they grow older and they cannot contribute to national economic performance to their true potential. Even in successful school systems such reasons direct attention to the performance of schools for all their students including those groups for whom, in the past, they were less successful.
- Decentralisation. For a variety of reasons many countries have begun to decentralise their educational systems and devolve increasing degrees of decision-making to lower levels in the system. This has led to greater diversity in schooling and less knowledge centrally of what is going on at institutional level. In devolved systems there is a developing feeling that the counterpart of greater freedom at institutional level is an increased need for accountability to show how such freedom is been used (OECD, 1995b).

Whilst these have operated to some degree in all educational systems, the priority accorded to particular reasons has varied as have the resulting policy decisions. A number of factors have influenced the outcomes in different countries, including the degree of satisfaction with the educational system, its status and the degree to which it is being used as a vehicle for social and economic change. In addition to these political issues, there may also be structural factors.

Levels and sectors affected

These issues have affected all levels and sectors of the educational system but the intensity has varied by country and by sector. Levels refers to levels in the administration of education – national, upper intermediate, lower intermediate, institution – whilst stages means the various age ranges of students in educational institutions – pre-primary, primary, lower secondary, upper secondary, post-compulsory, higher education (OECD, 1995a). Cor-

respondingly the particular evaluation machinery has differed by sector. In this chapter the major preoccupation will be with accountability and evaluation in the school sector since this is by far the largest sector and has generally been subjected to the greatest degree of accountability and evaluation in most countries (OECD, 1995b).

A general trend in educational reform has been for power to be concentrated at the level of central government and in institutions. Other levels have generally had their power and influence limited so the analysis of evaluation within the school system will concentrate on the effects on the centre and on institutions.

Increased attention to accountability and evaluation has not been confined to education. Power (1997) has drawn attention to the extent to which societies have begun to identify audit and evaluation as at least part of the solution to any problem. He refers to this trend as the 'audit society'.

Public and private sectors

Whilst the main emphasis here will be on public sector institutions – those wholly financed by public resources and controlled by public bodies – over the last 10 years there has been an increase in wholly or partially private institutions (OECD, 1995a). The wholly private are those financed and governed by private individuals or corporations. A large growth area has been of institutions that receive some public funding in addition to their private income. The state has generally regulated this sector in a variety of ways and this may include some element of evaluation. Although increased accountability in the public sector has raised awareness of such pressures on the private sector, governments have not generally intervened. Italy is an exception since it assesses private schools as part of the certification process whenever they apply to be acknowledged by the state (Hopes, 1997).

Purposes of evaluation

Institutional evaluation, like staff appraisal, is generally held to have two main purposes – accountability and development. Whilst these two ideas are linked, combining them in the same process has its tensions. Whilst accountability is intended to be made public, this is inimical to an honest assessment of performance, particularly where there may be deleterious consequences of not receiving a favourable report. Thus, particularly in a competitive situation, there is a great incentive to disguise or suppress any shortcomings and to appear to be performing well. The higher the stakes, the greater the incentive. This, however, is not consistent with the kind of assessment which would be needed as the basis for improvement (Fidler, 1997). This needs to be an accurate assessment of current performance so that any shortcomings can be

accurately diagnosed and a realistic improvement plan formulated. Thus an assessment for the purposes of improvement would need to be carried out privately so that there was no incentive to put on a show.

For either accountability or improvement the ideal is an accurate assessment of current strengths and weaknesses. This will be least likely where there are adverse consequences of a poor assessment and where the assessment is carried out in an adversarial spirit.

Much rhetoric appears to treat accountability as an end in itself rather than as a means to an end. Here it will be assumed that the desired ends involve policy-making and action at some level in the educational system which lead ultimately to improvement of some kind. Once the self-evident value of evaluation for accountability is questioned, its value changes with the type of assessment which is carried out, and the level within the system where the results are assessed. Further, the consequences of accountability depend on the type of accountability which is operating. These factors will be explaining in the following section of this chapter.

Evaluation for accountability

Although accountability is the most often cited reason for external evaluation, this often means little more than the assessment of institutional performance in the public domain. The potency of this rather depends on the consequences which follow from a performance assessed as poor. Where a small number of institutions are publicly pilloried for poor performance, then the impact may be substantial – not necessarily helpful, but substantial. However, where a large number of complicated and technical assessments are made public at the same time, the impact on each institution may be quite small. So some analysis is needed of mechanisms for accountability.

Kogan (1988: 25) takes as his definition of accountability 'a condition in which individual role holders are liable to review and the application of sanctions if their actions fail to satisfy those with whom they are in an accountability relationship'. He recognises that any definition of accountability may be contested but he differentiates accountability from more general forms of pressure and feelings of moral responsibility.

He accepts the analysis of Becher et al. (1979) that in schools a teacher has three kinds of accountability:

- to one's clients (moral accountability)
- to one's colleagues (professional accountability)
- to one's employers or political masters (contractual accountability).

However, he detected an emerging further type – market accountability. This was where clients have a choice of institution that they might attend. Whilst

the other forms of accountability could be seen to bear directly on an individual, this further form was mainly directed at the institution rather than an individual teacher. As Hirschman (1970) has suggested, there are two expressions of market accountability; 'exit' for those entering or leaving an institution and 'voice' for those who remain. The analysis of market accountability has tended to focus on the former since some degree of choice of institution is a hallmark of a competitive arena whereas working with existing clients, and having some form of accountability to them, is a more general feature of educational systems. If market forces are to be used to drive up standards and influence consumer choice, any evaluative information must be available to them in a form which they can comprehend.

Where the performance of an individual teacher is assessed the locus of accountability is relatively clear. I qualify the clarity because an assessment is needed as to whether the task which the teacher is performing in a particular context is one that a competent practitioner should be expected to demonstrate to an acceptable standard (Fidler and Atton, 1999). Where this condition is satisfied then the teacher is accountable for his or her own performance. However, when it is the institution which is the unit of assessment, and this is the general trend (OECD, 1995b), the situation is more complex.

When the institution is the unit of accountability an analysis is first needed of the sanctions which can be applied to the institution and then a separate analysis of the sanctions which can be applied *within* the institution, resulting from an adverse evaluation.

Professional and moral responsibility may bear heavily on individuals, be the most substantial motives to deliver a high-quality service to clients and be potent spurs to improvement. However, within the tight definition given by Kogan (1988), they do not meet the test of involving explicit sanctions if the performance of an individual is found wanting. Hence the analysis here will concentrate on forms of accountability which do meet Kogan's test. This should not be taken to mean that professional and moral responsibility of an individual for his or her performance are unimportant, only that they cannot be enforced if the performance does not meet expectations.

Accountability at the institutional level The following are accountabilities which could incur sanctions for the whole institution:

- statutory non-educational requirements – these include employment legislation and health and safety requirements
- statutory educational and financial requirements – these include curricular requirements and publication of student performance data
- statutory arrangements for the control and direction of the educational institution

- market accountability – this includes any consequences of clients choosing to study in a competitor institution.

Within the institution The main enforceable accountability operates on individual staff through their contract of employment. This contract takes account of the statutory arrangements for the control and direction of the institution.

Thus performance within the institution is overseen by the governance and leadership of the institution in the large number of educational systems where the institution has a corporate entity. In countries such as Germany, where the accountability is directly on the teachers (OECD, 1995b), school supervisors deal with such issues.

Evaluation for improvement

Evaluation is intended to report on the current performance of individuals and institutions and be a spur to improvement either in preparation for the evaluation or as a result of the evaluation. The basis of this may be different for each accountability system and for institutions which are evaluated at differing standards of performance.

A further issue on which systems differ is the provision of advice and support following evaluation. It is generally recognised that to improve following the diagnosis provided by an external evaluation, further advice and support will be needed. In some educational systems this is combined with evaluation, in others there is a separate support function whilst in others it is the responsibility of the institution to seek out and fund its own advice and support (OECD, 1995b).

Means of evaluation

The four principal means of formal evaluation are

- external inspection
- external audit or review
- performance indicators
- self-evaluation (OECD, 1995b).

These are not mutually exclusive and are actually likely to be interrelated. The value of performance indicators as the sole data or as adjuncts to other mechanisms should be noted. Observers generally see rather greater potential for performance indicators than has been achieved at the present time (OECD, 1995b; 1995c).

Which of these is the principal mechanism varies by sector within a particular country and between countries. The unit of accountability also changes

between countries. With much delegation of decision-making to school level there has been a corresponding concentration of evaluation and accountability at the level of the school. However, this depends upon the educational system of a particular country.

As a recent Organisation for Economic Co-operation and Development (OECD, 1995b) report suggests, the methods used for evaluation, and the consequences of evaluation, reflect the satisfaction of the country with education, its national culture and the esteem with which teachers are held.

External inspection This involves some external professionals visiting the institution to collect data. Usually this involves the observation of teaching performance. Inspection may involve explicit criteria for data collection and describe the various levels of performance against which teaching and other activities are to be judged. Whilst some educational systems demonstrate this more systematic approach, others rely on the professional judgement of inspectors using more implicit criteria, e.g. Germany. A case study of the English school inspection system is presented later in this chapter as other countries are watching this experiment with interest (OECD, 1995b).

External audit External audit or review involves external evaluation but acts mainly as a check on an internal evaluation of performance which has been made against previously formulated institutional priorities. Thus although the range of aspects to be evaluated may be common across institutions, the priority accorded to different areas varies by institution. The assessment which is made is against individual institutional priorities and what is being assessed is the degree to which internal evaluation has made a valid assessment of performance against these priorities. Later in this chapter, a case study of school review in Victoria, Australia, illustrates extensive use of performance indicators to inform an internal assessment by school staff which is validated by an external verifier.

Performance Indicators (PIs) For the purposes of evaluation of institutional performance Gray (1995), following Gray and Jesson (1991), has suggested the following requirements of PIs:

- central to process of teaching and learning
- cover significant parts of a school's activities
- reflect a range of competing educational priorities
- capable of being assessed
- allow meaningful comparisons to be made (over time and between schools)
- be capable of change by the school
- be few in number (three or four are suggested).

They suggest three areas:

- academic progress
- student satisfaction
- student-teacher relationship.

The measure of academic progress would be calculated using a value added methodology which examined the difference between the exam or test scores when the student leaves compared to the test scores when the student arrived (Willms, 1992). Those schools ensuring that their students make more progress than a typical school would be high added-value schools.

Performance indicators can be of two kinds – outcome and process. Evaluative performance indicators tend to be outcome indicators but for institutions to improve they need some indication of the processes by which other institutions achieved better results. A combination of outcome and process indicators may help identify other organisations for benchmarking (Fidler, 1999).

There are problems in trying to evaluate performance in multi-indicators systems unless the relative weightings of the various outcome indicators have been specified in advance. As Mayston and Jesson (1988) point out, improvements in one indicator may be achieved at the expense of another, so, unless the trade-offs are known or specified, higher performance may be ambiguous. A further issue to be considered in using performance indicators for institutional evaluation is that, as 'higher stakes' assessments are made, there is an increasing risk that those who provide the data will feel pressurised to massage or falsify the information as a form of institutional protection against adverse consequences. The risks of untrustworthy data should not be underestimated.

Self-evaluation This may have different forms and may be highly structured with areas to be assessed, and even have means suggested for collecting data to make the evaluation, but the judgement is an internal one. Issues raised concern the extent to which valid comparisons can be made by those intimately involved in the activity and the extent to which unpalatable judgements will be made. Beyond that there are issues of the ability and will to make changes, particularly where they may be radical and may have deleterious consequences for the teachers involved (Clift, 1982).

How the evidence is used

As Wilcox (1992) has noted, the means by which evidence is assessed to yield judgements is not well understood and yet it is critical for valid assessments. For improvement it is important both that any evaluation should be a valid

one and also that those within the institution accept the assessment as valid. This latter requirement is particularly difficult if the assessment is a negative one since there is an inevitable tendency for those involved to deny the results and to rationalise such denial as resulting from an assessment which is not valid (Earley, 1997).

The requirements for a valid assessment involve the following stages:

- (1) common, defined criteria for the collection of evidence and judgements
- (2) collection of representative and valid data
- (3) the comparison of these data with explicit norms for other schools
- (4) judgement about the meaning of differences emerging from the previous stages.

Comparison is the main mechanism by which informed judgements can be made. The different forms of comparison are:

- comparisons of one institution's data over time
- comparison with all institutions
- comparison with similar institutions
- examination of the profile of responses (Glogg and Fidler, 1990).

Difficulties arise at the third stage, particularly for qualitative judgements. The norms are likely to be implicit and be derived from the accumulated experience of the inspector. At a minimum this implies some requirements for prior experience to form these implicit norms. Both this and the final stage are judgemental and depend to differing degrees on personal opinion (Fidler et al., 1998).

There are well-known psychological biases in decision-making which arise from biased comparisons. These include:

- heightened recollection of the most recent experience
- high-profile examples dominate
- a search for consistency – good or bad (Fidler, 2002).

In the management literature (Hammond et al., 1998) these and others are well-known potential biases but they have not been widely discussed in education, nor have procedures been suggested to lessen them in the evaluation of institutions.

Evaluation at differing levels in educational systems

Evaluation can be considered at the following levels:

- system
- institution
- teacher.

The purpose of evaluation at different levels

At the system level, national data provide evidence of school standards and their change over time. Evaluation data can provide evidence on the progress of school reforms, the compliance of schools with national requirements and the responsiveness of schools to parental and societal demands. Such data facilitate an analysis of value for money and are part of general democratic accountability (OECD, 1995b). In a study of inspection in European Union countries, the tracking of school reforms was the least used outcome compared to institutional or individual evaluation (Hopes, 1997).

At the institutional level, evaluation identifies strengths and weaknesses which may aid school improvement. Schools with very poor performance are identified for support. Institutions may be encouraged to engage in self-evaluation in preparation for external evaluation and new ways of working may be encouraged. Student performance should benefit.

At the level of the individual teacher, evaluation identifies their strengths and weaknesses and may aid their improvement. External inspection can provide an independent assessment which can be compared to internal school judgements of teacher performance. Very poor teaching performance can be identified.

The contribution of different mechanisms of evaluation at different levels

System level External inspection can make a substantial contribution to knowledge of what is happening and the standards reached in each institution. This can be achieved either by inspection of each institution on a regular cycle or by inspecting a representative sample of institutions. Where all institutions are being inspected over an extended period, then the results for any particular year will depend on the findings from the institutions inspected in that year. If the sample is biased by the overrepresentation of certain types of institution then this will distort estimates of the performance of the system. For example, in the case of England, where in the first year of the new inspection regime those schools likely to have problems were scheduled for inspection in greater numbers than would have been representative, this gave an overly negative picture of school performance. Such schools were scheduled for an early inspection as a way of speeding up their recognition and improvement (Ofsted, 1999). In this case raising individual school performance was chosen as an overriding priority rather than achieving an accurate picture of system performance.

External review makes a more limited contribution to knowledge of the system because each institution is being judged according to its own mandate and priorities. Although these can be recorded as part of the review evidence they are more difficult to conflate and summarise at system level.

Performance indicators make a major contribution to knowledge of system performance but only on the areas covered by quantitative indicators.

Institutional self-evaluation plays almost no part in assessing system performance since the judgements are made by a large number of different individuals on differing bases and the people making the judgements have a vested interest in the result.

Institutional level The relative contributions change at the level of the individual institution. Whilst external inspection and performance indicators continue to make major contributions by comparing the performance of the institution with that of others, external audit and self-evaluation make greater contributions. External audit compares the performance of an institution with its own set priorities in a systematic way and gives it external legitimisation. This may involve some comparison with other institutions where there are common criteria.

Institutional self-evaluation engages staff in making an assessment of the performance of their institution. Although they may have a vested interest in the result, if the assessment involves comparison of data with other institutions it will be an assessment which takes account of prevailing standards of performance elsewhere. The range of areas of an institution's activities which are evaluated can be wider than other forms of evaluation and can be chosen because of institutional priorities for improvement.

Teachers Teachers have been included to demonstrate that, with the move to concentrate on institutional assessment, the performance of an individual teacher has become more peripheral to the evaluation process. Teachers may be assessed as part of external inspection but each teacher will be seen so little that any assessment will be very slight.

The performance of teachers can be inferred from examining the progress which their students make in a year when they have been tested at the start and end of the year but, as the number of students in a class is not large, the errors involved in making such an assessment mean that only exceptional performance – good and bad – stands out. The use of such statistics is also the only way in which the performance of individual students could be tracked.

Figure 18.1 summarises the impact of different forms of evaluation.

Case studies

Two case studies give details of two very different approaches to inspecting schools. One uses external inspection as the principal mechanism and operates

Level	Ext inspection	Ext audit	Pls	Inst self-evaluation
System	++	+	++	0
Institution	++	++	+	+
Teacher	++	0	+	+

Figure 18.1 The contribution of different mechanisms of evaluation at different levels

in a competitive school environment – this is the case of England. The other uses external verification of an internal audit using a large number of performance indicators – this is the case of the state of Victoria in Australia.

School inspections in England under the Office for Standards in Education

This is a scheme to inspect all state schools which began in 1993. Over four years all 23,500 state schools have been inspected. Major modifications have been made to the scheme on a number of occasions and a number of the original criticisms have been met – data collection before the inspection has been streamlined, brief feedback to teachers after they have been observed has been instituted, developments on which schools are already working are recognised, briefer inspections have replaced standard inspections for those schools which appear to be making good progress, ongoing training has been provided for inspectors, inspectors with a poor record have been dismissed and an appeals procedure has been put in place. The changes make it quite difficult to provide research evidence on the working of the scheme because some of the research evidence becomes outdated when the scheme is modified. However, where it is appropriate, evidence will be cited.

The particular features of this inspection scheme include the following:

- systematic external inspection of all schools on a six-yearly cycle
- separates advice and inspection
- published criteria for inspection – content and process
- reports are made public
- high-stakes consequences for poorly performing schools
- independent inspectors operating under contract
- presence of a lay inspector
- comprehensive account of a school's performance
- systematic classroom observation and aggregated assessment
- action plan required from school afterwards.

The scheme is administered by a specially created Office for Standards in Education (Ofsted). In principle this is independent of the central government education department. This office organises but does not conduct the inspections. It created an explicit set of criteria and set of protocols for inspections

and these were published and sent to schools. Thus for the first time schools were aware of the inspection criteria and arrangements and could prepare for inspection.

The inspections are carried out by inspectors who have been accredited by Ofsted following a short period of training and assessment. The inspection of a school is carried out by contract. A range of organisations tender for inspection contracts and of those which have passed a quality threshold; the lowest cost tender is accepted. The successful contractor is then responsible for organising the inspection and submitting a report afterwards. Contractors include LEAs, non-profit-making organisations and commercial firms.

All except the lay member of the inspection group are expected to have relevant professional experience in schools and to have appropriate subject expertise. Research has shown that this is quite difficult to ensure. Initially there was a shortage of inspectors with primary teaching experience and, because of the scheduling of inspections, inspectors are brought into teams at short notice and may not have the exact subject match for the role they are undertaking. Many inspectors do not have recent teaching experience and a number are retired teachers (Ferguson et al., 2000). A small number of inspectors are full-time teachers or headteachers in other schools. This can be particularly developmental for those individuals who report that they take a more analytical view of teaching in their own school and also are able to see the detailed working of other schools that they are inspecting.

The new inspection regime was intended to divorce inspection from advice and involve a fresh examination of a school. Thus inspectors identify strengths and weaknesses in teaching and the general operation of a school but do not give direct advice on how to improve and have no continuing relationship with the school after the inspection (Ferguson et al., 2000). Some schools have taken the opportunity to employ inspectors afterwards to help them make changes, but that is a school decision and a school expense. Inspectors were also intended to have no recent prior knowledge of the school they were inspecting so that their evaluation would be unbiased by any previous experience at the school. This has proved quite difficult to operationalise when it is contractors who organise the inspection and may need to make changes to planned inspection teams through the unavailability of some members (Ouston et al., 1996). The inspection of some 5,000 schools per year, when these inspections have to take place in a limited number of weeks in the school year, is a huge logistical operation.

The inspection involves a number of inspectors visiting a school over a number of days to observe teachers teaching, to sample student work and to interview other interested parties. There is a simple parental satisfaction questionnaire which is sent to all parents and there is a meeting of parents held by the lead inspector at which parents can air their views of the school. The inspection group have previously had access to a great deal of

documentation from the school and its performance indicators. For a small primary school three inspectors might spend three or four days inspecting, whilst for a large secondary school up to 10 inspectors might spend up to a week in the school.

The inspection team are expected to sample between 10 and 20 per cent of the lessons taught in the period they are in the school, although this may involve only partial lessons. They are expected to grade each teacher on the teaching that has been observed and give brief feedback. Both inspectors and teachers report dissatisfaction with the brevity of the feedback and the pressured circumstances in which it is conducted (Ferguson et al., 2000). All these data from observation and all other sources are systematically aggregated and used as the basis for making qualitative judgements about the performance of the school and its teachers. All these data, typically comprising some 2,000 pages for each inspection, are stored on a huge database overseen by Ofsted and which is used to make authoritative statements about the working of the educational system.

The inspection team are required to make corporate judgements on four areas of a school's work:

- the quality of education provided by the school
- the educational standards achieved by the students
- the efficiency with which resources are managed
- the spiritual, moral, social and cultural development of the school.

The results of the inspection, including strengths, weaknesses and areas for improvement, are fed back to senior school staff at the conclusion of the inspection and a full written public report is issued within six weeks. This is available on the Internet, is distributed to newspapers and other organs of publicity, and a summary is sent to each parent at the school.

As part of the inspection regime a group of schools was defined for the first time. The worst schools (about 2 per cent of all schools) are deemed 'in need of special measures' or, more colloquially, 'failing schools'. These are schools which have a range of weaknesses. Usually these weaknesses involve poor examination or test results, a low proportion of satisfactory teaching, poor student attendance and other weaknesses (Ofsted, 1999). Whilst these schools are disproportionately in deprived inner city areas, this is not always the case. Some schools have failed because children were not making sufficient progress and the teaching was of too low a standard (Ferguson et al., 2000).

More typical schools are required to formulate an action plan for improvement, but there is no systematic monitoring of progress on this plan until the school's next inspection. The original design was for inspections every four years but this has been lengthened to every six years for

typical schools, although schools causing concern may be inspected more frequently.

Since the year 2000 about 30 per cent of schools which have good performance indicators have shorter inspections where not all staff are observed teaching. However, expectations for these schools are very high and so shorter inspections do not appear to have lessened the stress on teachers which inspection has induced.

The publication of inspection reports has been part of an experiment with the whole school system in England (OECD, 1995b). That inspection has been accepted as readily as it has, is in no small part due to the experience and judgement of the initial group of inspectors who had to operate the system and the explicit criteria on which the inspection was to be conducted. There have been many criticisms of the new regime. A trenchant one concerns the lack of demonstrated validity of inspector judgements following observation of teaching (Fitz-Gibbon, 1997). A voluntary comparison of grades between inspectors has demonstrated a fair measure of agreement but the use of volunteers almost certainly makes the result unrepresentative (Mátthews et al., 1998).

Undoubtedly this form of inspection has engendered a great deal of stress in teachers and inspectors (Jeffrey and Woods, 1998). A more insidious concern is the extent to which the inspection criteria dominates teachers' discourse and set up an orthodoxy (Maw, 1996). However, a House of Commons (1999) investigating committee was not persuaded that this was a sufficient danger to suggest major changes to the inspection regime. They noted that headteachers in surveys have reported that inspection has been to varying degrees developmental for their school and led to improvement (Ouston et al., 1996).

Public knowledge of the working of the educational system is much greater and more soundly based than previously. This is aided by external public testing and end of school examinations at 16, which apply comparable standards to each school and from year to year. However, improving performance indicators and improving inspection results from schools as a whole do not seem to have lessened politicians' concerns about the performance of schools.

The small number of poor schools are now public knowledge and action is being taken, often in very difficult circumstances. Informed professional opinion considers that these schools have been in this state for some time and it is the new inspection regime which has exposed them rather than created them (Watling et al., 1998).

Finally, it should be remarked that increased research interest in inspection brings effects to light which would otherwise be unknown. It would be surprising if similar effects are not operating in other inspection regimes.

School review in Victoria, Australia

A very different approach to school review has been taken in the state of Victoria in Australia. This framework makes far greater use of performance indicators, including a good deal of data specially collected by each school, and examines each school's progress against its charter of priorities. The scheme began in 1996 and all 1,600 state schools have been reviewed over a three-year cycle (Ferguson et al., 2000). There are two main purposes of the framework: 'to satisfy legitimate expectations of government about accountability for the outcomes of schooling, and to assist schools and teachers to improve standards of student learning' (Office of Review, 1998: 4).

Particular features of this scheme are:

- audited self-evaluation of progress against an individual school charter every three years
- presence of an external verifier
- extensive use of comparative data
- school data used in addition to specially collected data – some mandatory and some at the school's discretion
- specially written computer software to facilitate school comparisons of data over three years
- analysis of student level data
- staff involvement in interpreting data
- results incorporated in the next school charter.

General inspections of schools had ceased some 20 years previously and so, in devising this new system, much attention was paid to what was happening in other countries and the following were some guiding principles which influenced the design:

- external evaluation is more effective in improving school performance when schools have well developed internal processes
- school self-evaluation without some external component lacks the rigour necessary to effect real and lasting improvements in school performance
- evaluation processes should assist schools not only to analyse their performance but also to improve the effectiveness and efficiency of their management practices (Office of Review, 1998: 4).

This is a regime overseen by a specially created section of the State Education Department, the Office of Review (the Accountability and Development Division since 1999). It set up the system and oversees its operation, and is a good deal simpler and less costly than Ofsted. There are no inspectors.

Schools review themselves by comparing their collected data against that of other schools, particularly 'like schools'. These are schools within a similar band of free meals and proportion of children whose first language is not English.

The sole external contribution to the review is the presence of a verifier at a review meeting at the school towards the conclusion of the process. Ten organisations have verifier contracts and supply a person to act as verifier to each school review. The task of the verifier is to ensure that data have been collected appropriately and that the results have been professionally considered by teachers at the school for their implications for school performance and, finally, that improvement targets emerge from the process.

The whole process revolves around a school's charter. This is the set of priorities which are set up by the school in conjunction with its parents for the next three years. Each year for the following two years the school collects data on a range of criteria and compares its results with state averages and with those from like schools. Either this process, or new state mandates introduced during the year, provide developmental targets for the following year. In the third year there is a cumulative examination of the data and appropriate comparisons. The results of the review are expected to inform the school's next three-year charter (Gurr, 1999a).

The range of areas on which data are collected fall into two groups. First, there are data which are collected routinely by a school as part of its record-keeping on student and staff activities. Second, there are data which are specially collected by surveys for the purposes of review:

- parental satisfaction
- staff satisfaction.

The data are stored on computer and specially written software facilitates the display of data for comparison purposes.

As these data are presented and compared with other schools, a substantial number of inferences and questions arise. These include questions about classroom teaching even though there has been no classroom observation. An approach such as this can identify areas in need of further scrutiny in a very discriminating way.

In Victoria market mechanisms are weaker than in England and, so, in addition to any professional and moral responsibility to improve, it is the bureaucratic accountability of schools to the central office which is a sanction which could be used if schools did not take the self-review seriously and formulate challenging improvement targets. The verifier role is a crucial one in assessing the extent to which this has been done. A dissatisfied verifier could write a note of dissent following the review process which would be sent to the central office.

Some evaluative data on the working of the accountability framework is available from a report by the Victorian Auditor General's Office (1997) and one commissioned by the Office of Review from verification panel members (Gurr, 1999b). These indicate an improvement in planning and reporting by schools and more active involvement by schools in the review process:

the framework presents an integrated programme that works at two levels: school planning and development, and system accountability. It is this dual utility that is the key to success. Schools value the framework for it provides them with a helpful developmental tool . . . the Victorian Accountability Framework is primarily an integrated school development framework, but one which also meets government accountability needs. (Gurr, 1999b: 10)

Whilst I consider that the assessment of school involvement in the improvement process can be substantiated, I am rather less convinced about the extent to which system-wide performance is evaluated. The centrally collected data provide some benchmarks against which progress in state schools can be assessed – attendance, examination results and student destinations after school. However, the only state-wide measure of student progress is the school-leaving examination at 18, although there are voluntary tests at other ages. Most of the data on student performance is provided by individual teachers in schools and with little cross-moderation. However, there are process data on the curriculum covered and much attitudinal data.

The curriculum data do provide evidence of the effectiveness of policy developments. For example, recent curriculum changes have required a 'Language other than English' (LOTE) to be taught in each school. Curriculum reviews have identified which languages have been chosen including the reasons. This has revealed the availability of language teachers to be an issue for primary schools.

The results of the review are available to parents in each school but are not publicly available. Thus they play a greater part in 'voice' rather than 'exit' (Hirschman, 1970). The main pressure for improvement comes from the state education department and its local offices.

Conclusion

The reasons for evaluating institutions need to be clearly specified. These should influence the choice of evaluation method and how it operates. Combining evaluation for accountability and evaluation for improvement involves tensions and compromises. There is also a further tension between system-level requirements and those of individual institutions. These tensions

are illustrated in the two case studies and the design choices which were made in setting them up.

A systematic process which covers all schools is needed both to enable valid statements about the performance of the system and so that poor schools can be identified. Doing this once provides a sound benchmark against which any future data can be compared. The two case studies illustrate different approaches to this systematic process.

Educational systems without such a regime can have no certainty about the number and location of schools with very poor performance. However, identifying such schools is only the precursor of a very difficult improvement process. Such schools often have a long history of poor performance which has become institutionalised, and such organisations are not easy to change. The introduction of the new inspection regime in England was used as an opportunity to set up procedures to tackle 'failing' schools as they were discovered, whilst in Victoria the existence of this group of schools has emerged as a result of the review process.

Valid inspection requires clear published criteria so that the evidence base and the basis for judgements are known both to inspectors and those inspected. However, running any evaluation regime for too long unchanged will cause it to lose impact because schools operate to the expectations of that evaluation regime. There is some evidence of Ofsted criteria being used in school self-reviews and hence reinforcing an orthodoxy.

The collection of data for performance indicators by schools has a number of advantages. The data are more neutral than judgements presented by inspectors, and teachers can be expected to accept the data and make their own inferences about their meaning. An extensive range of indicators is needed to give a rounded picture of each institution's performance. The Australian experience demonstrates both the extent to which data can give a comprehensive view of a school's performance and also the degree to which staff accept and examine the implications of the data.

If the assessment is perceived as 'high stakes', steps are needed to ensure that the data on which the evaluation is being made are valid and accurate since there will be increasing pressure on schools to massage the data to avoid adverse consequences for the institution. There are some indications that this is happening in England because the data are made public and may have serious consequences for a school.

An external dimension to this process is needed both to ensure that valid data are being collected and used and that appropriate inferences are being made. Using performance indicators could have the effect of familiarising teachers with the value of data and making them more reflective about using evidence to guide their practices. The English and Australian experiences demonstrate different ways of achieving this. In Victoria, a key role is played by the verifier as the sole external participant in the process.

Whilst performance indicators can be of value to good and typical schools, poor schools exposed by them need further help and support to improve. It is probably in such cases that teams of inspectors visiting the school can help confirm in greater detail the nature of strengths and weaknesses and help formulate priorities for improvement.

Further reading

- Earley, P. (ed.) (1997) *School Improvement after Inspection? School and LEA Responses*, London: Paul Chapman Publishing.
- Ferguson, N., Earley, P., Fidler, B. and Ouston, J. (2000) *Improving Schools and Inspection: The Self-Inspecting School*, London: Paul Chapman Publishing.
- Organisation for Economic Co-operation and Development. (OECD) (1995b) *Schools under Scrutiny*, Paris: OECD.

References

- Becher, T., Eraut, M., Barton, J., Canning, T. and Knight, J. (1979) 'Accountability in the middle years of schooling: an analysis of policy options', final report of the East Sussex LEA/University of Sussex Research Project, Brighton: University of Sussex.
- Clift, P. (1982) 'LEA schemes for school self-evaluation: a critique', *Educational Research*, 24 (4): 262-71.
- Earley, P. (1997) 'External inspections, "Failing Schools" and the role of governing bodies', *School Leadership and Management*, 17 (3): 387-400.
- Ferguson, N., Earley, P., Fidler, B. and Ouston, J. (2000) *Improving Schools and Inspection: The Self-Inspecting School*, London: Paul Chapman Publishing.
- Fidler, B. (1997) 'The school as a whole: improvement and planned change', in B. Fidler, S. Russell and T. Simkins (eds), *Choices for Self-Managing Schools: Autonomy and Accountability*, London: Paul Chapman Publishing.
- Fidler, B. (1999) 'Benchmarking and strategy', in J. Grant (ed.), *Sharing Experiences: Value for Money in School Management*, York: Funding Agency for Schools.
- Fidler, B. (2002) 'Strategic leadership and cognition', in K. Leithwood and P. Hallinger (eds), *Second International Handbook on Educational Leadership and Administration*, Dordrecht: Kluwer Academic.
- Fidler, B. and Atton, T. (1999) *Poorly Performing Teachers in Schools and How to Manage Them*, London: Routledge.
- Fidler, B., Earley, P., Ouston, J. and Davies, J. (1998) 'Teacher gradings and Ofsted inspection: help or hindrance as a management tool?', *School Leadership and Management*, 18 (2): 257-70.
- Fitz-Gibbon, C. T. (1997) 'Ofsted's methodology', in M. Duffy (ed.), *A Better System of Inspection?*, Hexham: OfSTIN.
- Glogg, M. and Fidler, B. (1990) 'Using examination results as performance indicators in secondary schools', *Educational Management and Administration*, 18 (4): 38-48.

- Gray, J. (1995) 'The quality of schooling: frameworks for judgement', in J. Gray and B. Wilcox (eds), *'Good School, Bad School': Evaluating Performance and Encouraging Improvement*, Buckingham: Open University Press.
- Gray, J. and Jesson, D. (1991) 'The negotiation and construction of performance indicators', *Evaluation and Research in Education*, 4 (2): 93-108.
- Gurr, D. (1999a) *From Supervision to Quality Assurance: The Case of the State of Victoria (Australia)*, Paris: International Institute for Educational Planning.
- Gurr, D. (1999b) *School Review Evaluation*, Melbourne: Office of Review.
- Hammond, J. S., Keeney, R. L. and Raiffa, H. (1998) 'The hidden traps in decision making', *Harvard Business Review*, September-October: 47-58.
- Hirschman, A. O. (1970) *Exit, Voice and Loyalty: Response to Decline in Firms, Organizations and States*, Boston, MA: Harvard University Press.
- Hopes, C. (1997) *Assessing, Evaluating and Assuring Quality in Schools in the European Union*, Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- House of Commons (1999) *The work of OFSTED*, Education and Employment Committee (Fourth Report of the Committee, Session 1998-99), London: The Stationery Office.
- Jeffrey, B. and Woods, P. (1998) *Testing Teachers: The Effects of School Inspections on Primary Teachers*, London: Falmer.
- Kogan, M. (1988) *Education Accountability: An Analytical Overview*, 2nd edn, London: Hutchinson.
- Matthews, P., Holmes, J. R., Vickers, P. and Corporaal, B. (1998) 'Aspects of the reliability and validity of school inspection judgements of teaching quality', *Educational Research and Evaluation*, 4 (2): 167-88.
- Maw, J. (1996) 'The handbook for the inspection of schools: models, outcomes and effects', in J. Ouston, P. Earley and B. Fidler (eds), *OFSTED Inspections: The Early Experience*, London: David Fulton.
- Mayston, D. and Jesson, D. (1988) 'Developing models of educational accountability', *Oxford Review of Education*, 14 (3): 321-39.
- Office for Standards in Education (Ofsted) (1999) *Lessons Learned from Special Measures*, London: Ofsted.
- Office of Review (1998) 'Building high performance schools: an approach to school improvement', paper presented at the national seminar on School Review and Accountability, Hobart, Australia, April.
- Organisation for Economic Co-operation and Development (OECD) (1995a) *Decision-Making in 14 OECD Education Systems*, Paris: OECD.
- Organisation for Economic Co-operation and Development (OECD) (1995b) *Schools Under Scrutiny*, Paris: OECD.
- Organisation for Economic Co-operation and Development (OECD) (1995c) *Measuring the Quality of Schools*, Paris: OECD.
- Ouston, J., Fidler, B. and Earley, P. (1996) 'Secondary schools' responses to OFSTED: improvement through inspections?', in J. Ouston, P. Earley and B. Fidler (eds), *OFSTED Inspections: The Early Experience*, London: David Fulton.
- Power, M. (1997) *The Audit Society*, Oxford: Oxford University Press.

The Principles and Practice of Educational Management

- Victorian Auditor-General's Office (1997) 'Schools of the future: valuing accountability', Special Report No. 52, December, Melbourne, Australia.
- Watling, R., Hopkins, D., Harris, A. and Berisford, J. (1998) 'Between the devil and the deep blue sea? Implications for school and LEA development following an accelerated inspection programme', in L. Stoll and K. Myers (eds), *No Quick Fixes: Perspectives on Schools in Difficulty*, London: Falmer.
- Wilcox, B. (1992) *Time-Constrained Evaluation*, London: Routledge.
- Willms, J. D. (1992) *Monitoring School Performance: A Guide for Educators*, Lewes: Falmer Press.



*U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)*



NOTICE

Reproduction Basis

- ☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
- ☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").